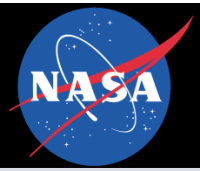


Langley Research Center

Creating “Intelligent” Climate Model Ensemble Averages Using a Process-Based Framework

Noël C. Baker and Patrick C. Taylor

Project funded through NASA Postdoctoral Program

April 23, 2014

CMIP5 climate models predict changes in TOA radiation during 21st century: high uncertainty in projections

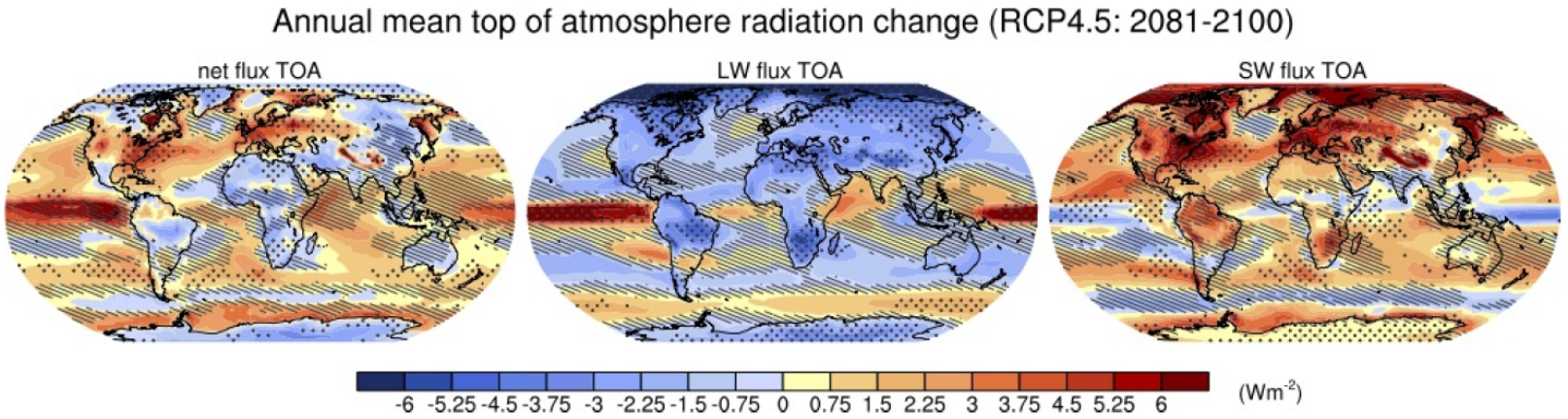
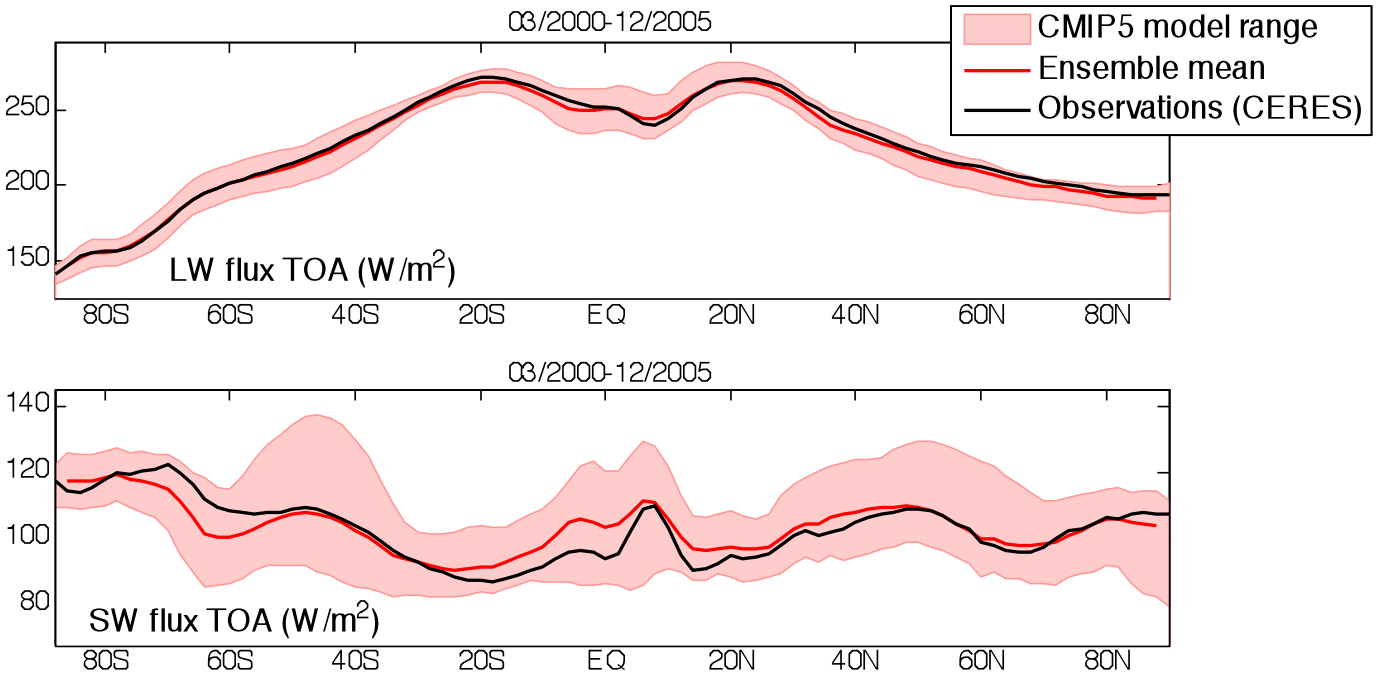


Figure 12.16 (IPCC AR5)
Anomalies are positive downward (relative to 1900-1950 base period).

Hatching = high uncertainty; stippling = low uncertainty and where 90% of models agree on the sign of change.

Large spread among model simulations, biases evident when compared with observations



- Individual models are subject to biases created by structural model uncertainties, and some models reproduce certain climate processes better than others.
- Ensemble averaging of multiple models is used to add value to individual model projections by constructing a consensus projection.

Ensemble-mean projections assume that models sample full range of possibilities

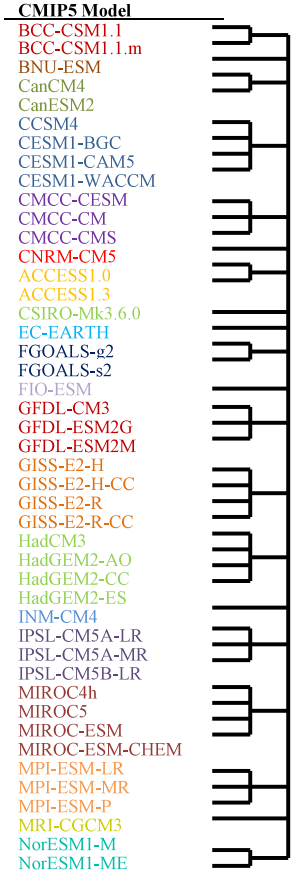
CMIP5 Model
BCC-CSM1.1
BCC-CSM1.1.m
BNU-ESM
CanCM4
CanESM2
CCSM4
CESM1-BGC
CESM1-CAM5
CESM1-WACCM
CMCC-CESM
CMCC-CM
CMCC-CMS
CNRM-CM5
ACCESS1.0
ACCESS1.3
CSIRO-Mk3.6.0
EC-EARTH
FGOALS-g2
FGOALS-s2
FIO-ESM
GFDL-CM3
GFDL-ESM2G
GFDL-ESM2M
GISS-E2-H
GISS-E2-H-CC
GISS-E2-R
GISS-E2-R-CC
HadCM3
HadGEM2-AO
HadGEM2-CC
HadGEM2-ES
INM-CM4
IPSL-CM5A-LR
IPSL-CM5A-MR
IPSL-CM5B-LR
MIROC4h
MIROC5
MIROC-ESM
MIROC-ESM-CHEM
MPI-ESM-LR
MPI-ESM-MR
MPI-ESM-P
MRI-CGCM3
NorESM1-M
NorESM1-ME

CMIP5: contains simulation data from ~45 models

Standard sets of experiments (run using same climate scenario forcings)

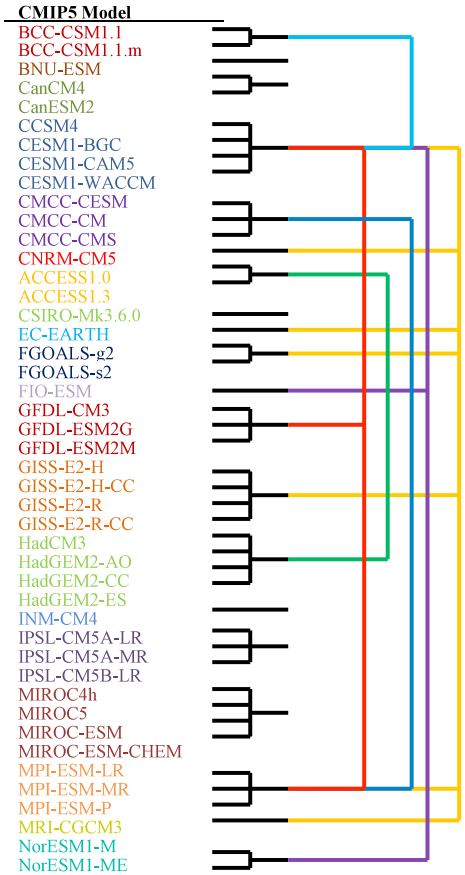
Typically ensemble-averaged to produce climate projections

Independent samples?



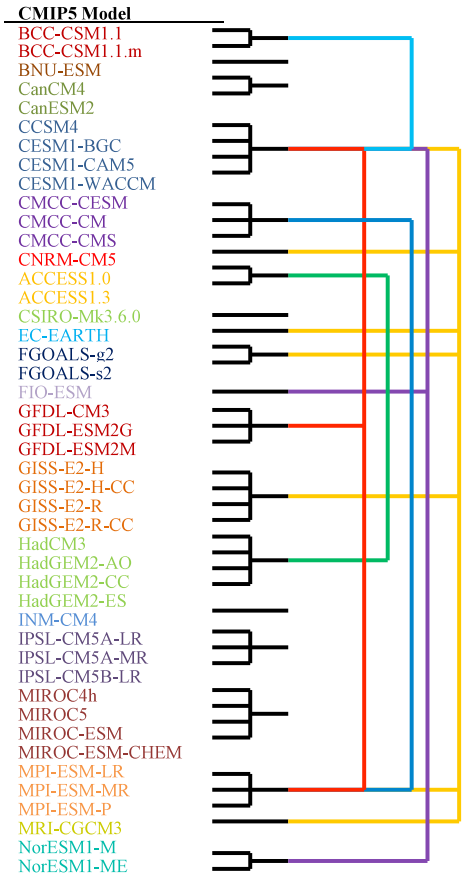
20 modeling groups from around the world:
each family shares
basic model components

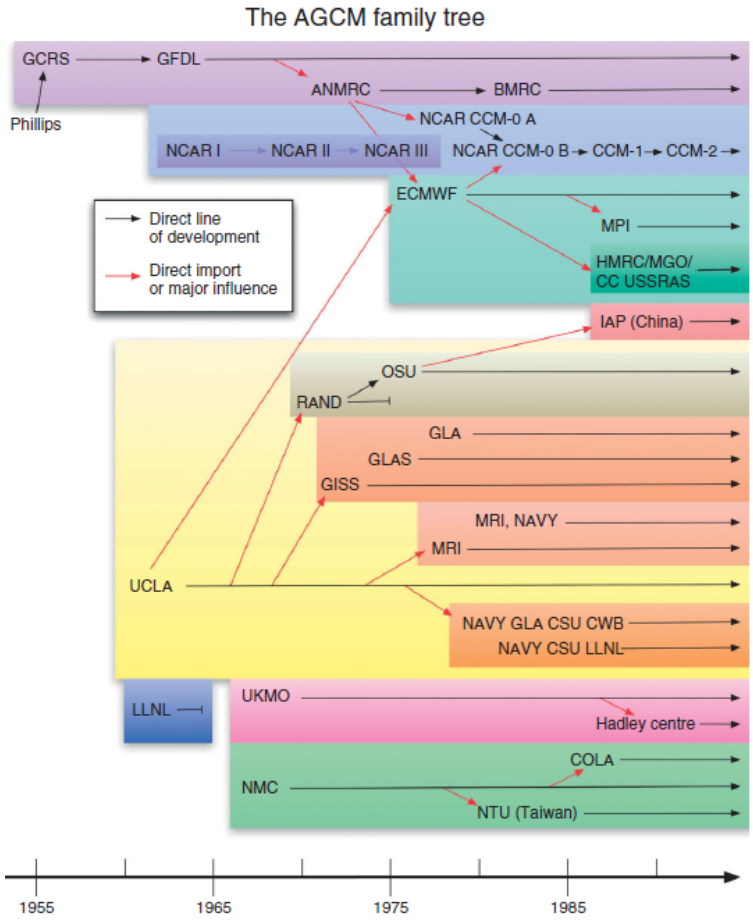
Models are not independent: do not sample full range of possible outcomes



Some families also share components or code, evident in performance (Knutti 2011)

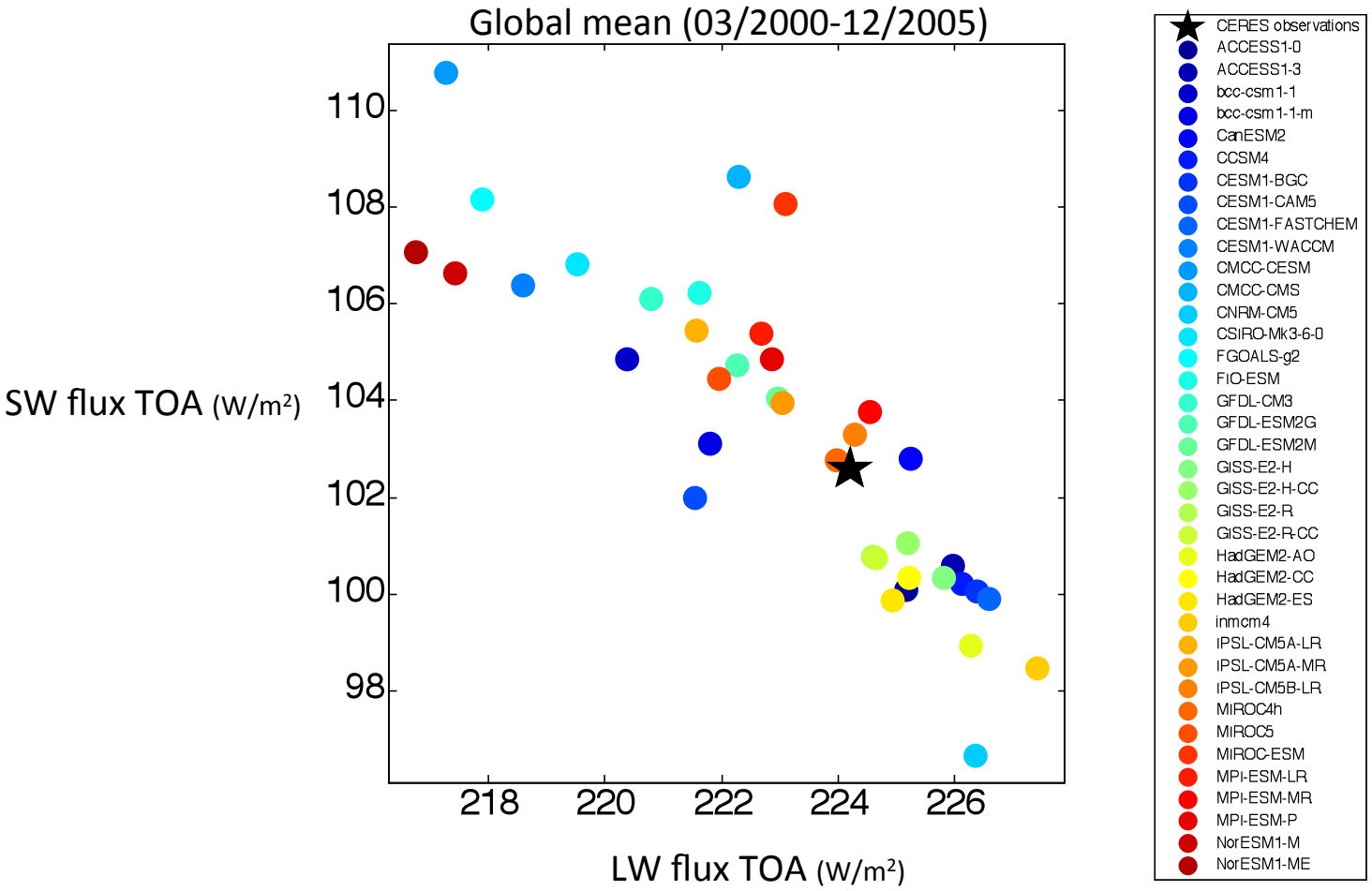
Models are not independent: do not sample full range of possible outcomes





Burnham (2010)

Some models perform better than others: is an equal-weight average the best method for combining model output?



Project questions:

Can we use knowledge of current climate to produce better future predictions?

Can we reduce prediction uncertainty by knowing which models perform better?

Is there a better way to make predictions than an equal-weight average of models of varying quality?

Possible solution: combine model output using unequal-weight ("intelligent") ensemble averages

Using knowledge of model performance, create and test metrics for producing "intelligent" climate predictions

Previous work has explored model performance and some unequal-weighting metrics

Several examples:

- Use only subsets of models (USGCRP 2009)
- Create mean-state metrics using model skill (Giorgi and Mearns 2002, 2003; Reichler and Kim 2008)
- Constrain model projections using mean-state CERES data (Tett et al. 2013)
- Weight using regression between observed and future trends (Boe et al 2009)
- Apply bias correction for present-day to future trends (Baker and Huang 2012)

"The community would benefit from a larger set of proposed methods and metrics" (Knutti 2010)

Project goal: design a framework for creating and testing new weighting metrics

How is this project unique?

- Framework will test:

1) A wide variety of metrics
2) Scale-dependence of metrics (global, regional, or gridpoint-scale weights)
3) State-dependence of metrics (do the weights change when created using different model experiments?)
4) Creating new ensemble-averaged projections of many different climate variables

- New, process-based metrics

- A process-based metric is defined as a metric based on the relationship between physically related climate variables

Example: ratio of outgoing longwave radiation to surface temperature (longwave component of climate sensitivity)

$$\frac{OLR\ anomaly}{surface\ temperature\ anomaly} = \frac{\delta OLR}{\delta Ts}$$

$$\frac{(longwave\ cloud\ component)\ anomaly}{surface\ temperature\ anomaly} = \frac{\delta LWcf}{\delta Ts}$$

Will also use a variety of process-based metrics (e.g. Nino 3.4 index, MJO)

Framework will be used to test important climate processes as model weighting metrics

Framework can test key science questions about climate model quality

- How well do models reproduce observed:

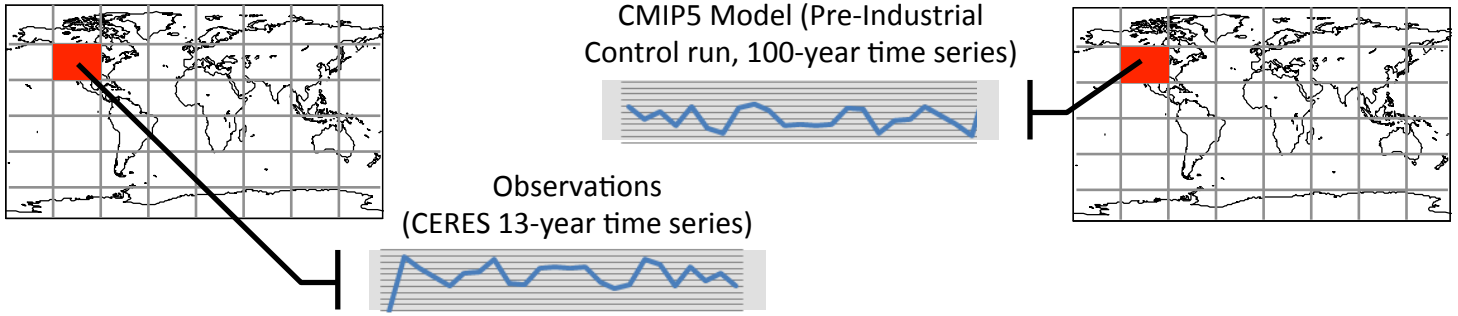
1) mean state
2) variability
3) frequency distribution

   for different tested quantities?

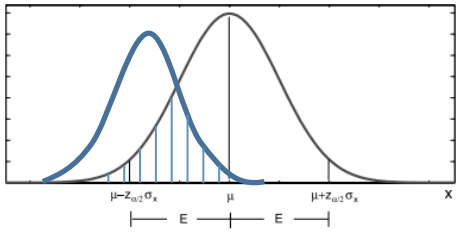
- What is the scale dependence of model quality? How does it vary when weight is computed globally vs. regionally (weight per grid point, per region, apply regional weights to globe, or one global weight)?

Example metric: weight models by how well they reproduce OLR frequency distribution

1) Select resolution (per grid point, regional, global)
2) Calculate OLR anomaly time series for the model (control run: Pre-Industrial Control) and observations (detrended)



3) Use statistical test to compare time series (Kolmogorov-Smirnov test for distribution similarity)
4) Calculate p-value of test to determine "amount of overlap" between distributions (large p-value: high similarity): $0 \leq p \leq 1$



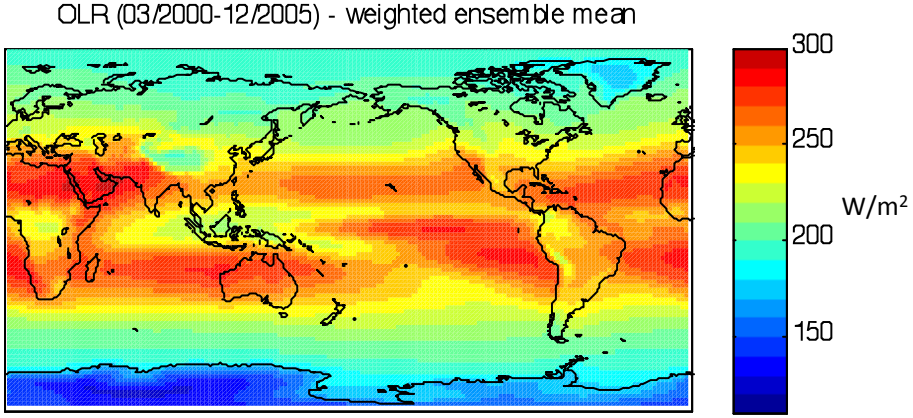
Example metric: weight models by how well they reproduce OLR frequency distribution

5) Repeat for each grid point: obtain map of weights for each model

.1	.6	.7	.3	.9	.5	.1	.4
.2	.4	.1	.6	.2	.3	.1	.5
.5	.2	.3	.2	.7	.6	.3	.5
.6	.7	.7	.8	.7	.8	.3	.4
.4	.3	.4	.7	.6	.2	.1	.4
.6	.4	.6	.4	.5	.8	.7	.3

6) Apply weights to chosen experiment (AMIP, 'historical', RCP future scenarios) for any variable (in this example: OLR, 'historical')

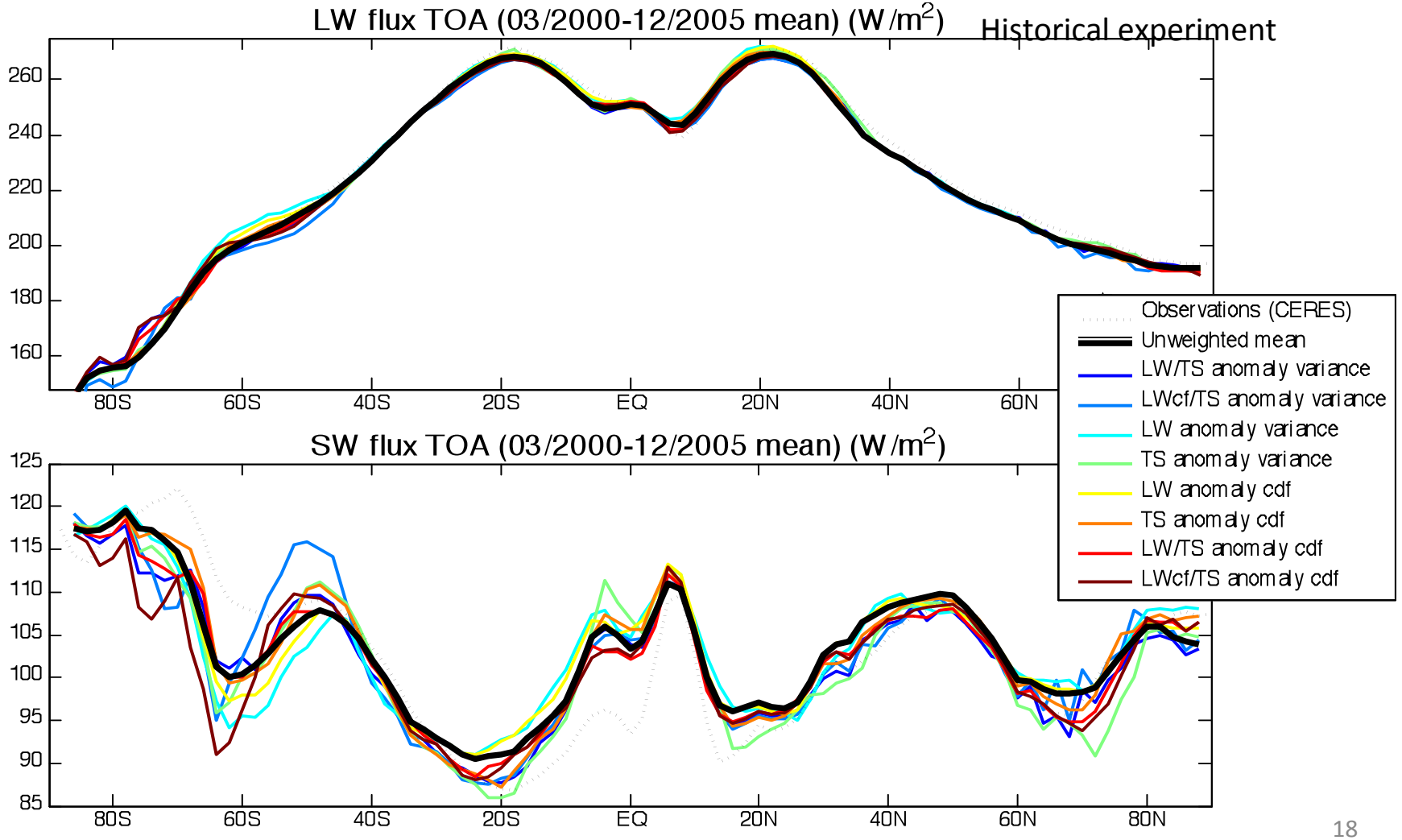
7) Calculate weighted ensemble mean



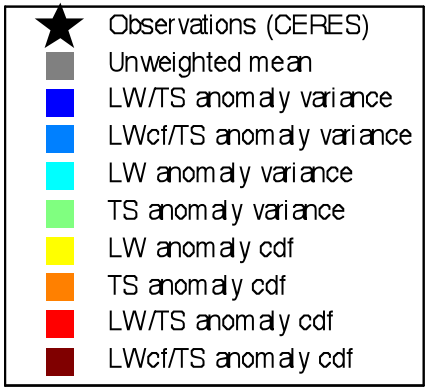
Project status

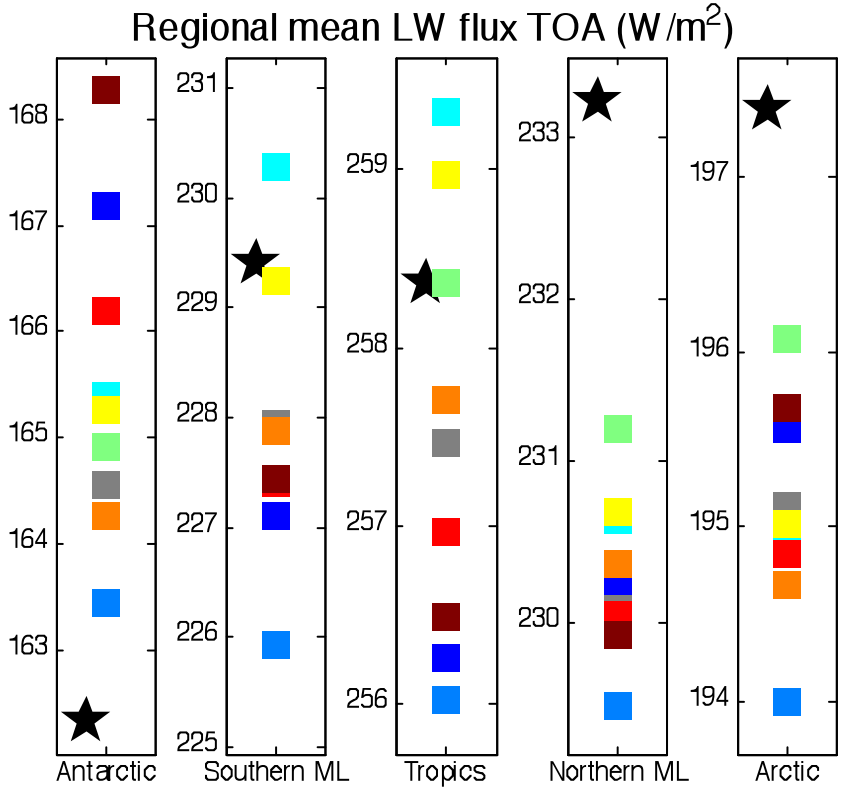
- Created 10 weighting metrics
- Tested on several quantities (LW flux, SW flux, precipitation, temperature anomaly)
- Tested on 2 different CMIP5 ensemble scenarios (present-day experiments): 'historical' (fully-coupled AOGCM simulations), AMIP (atmospheric model component only: driven by observed SSTs)

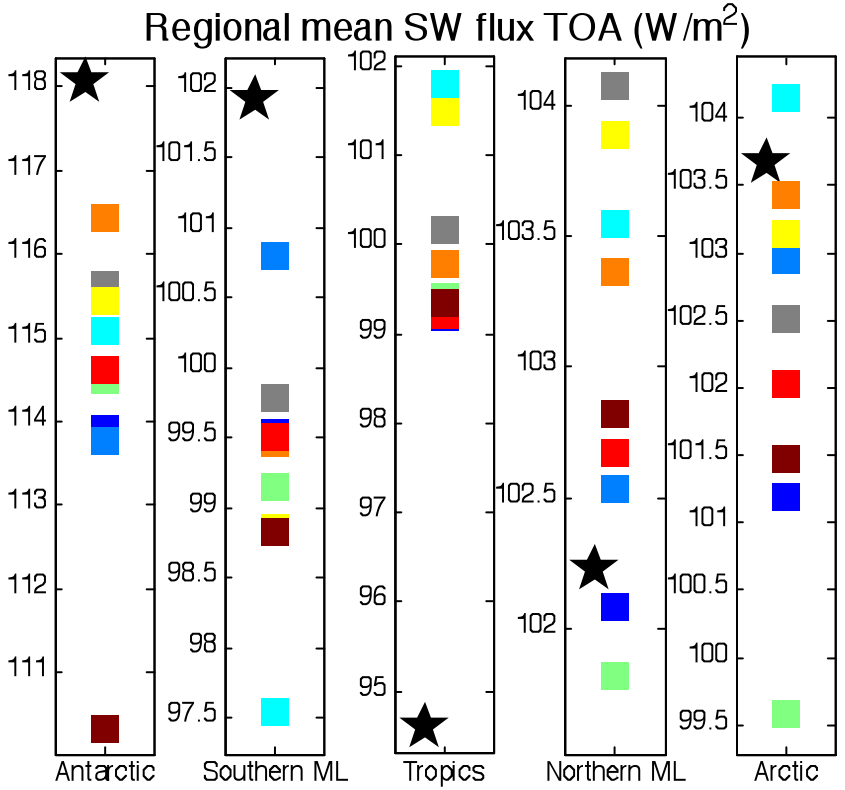
Initial results: weighted means can show significant differences from equal weight mean



Some weighting metrics improve model mean for certain variables and regions

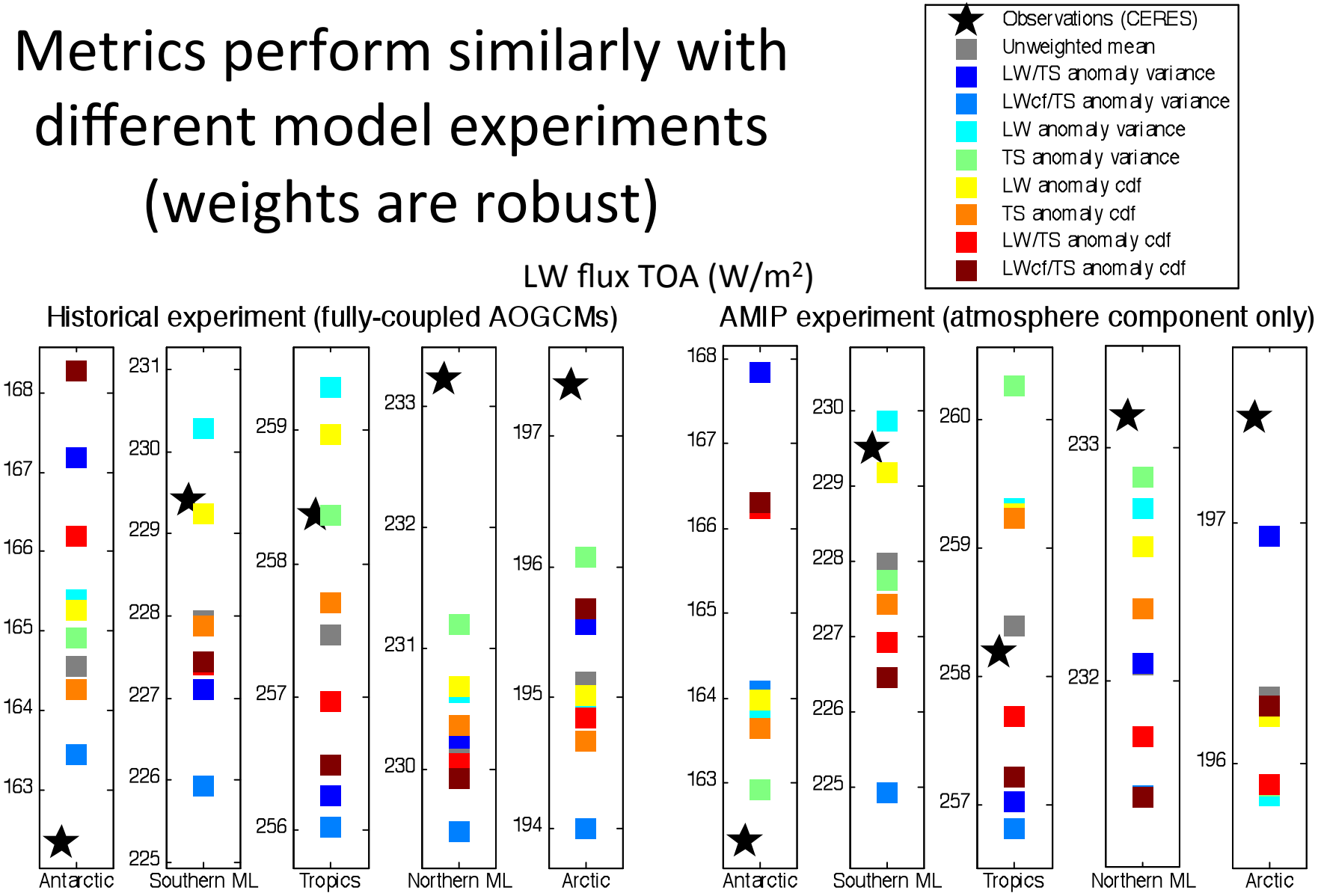






CMIP5 models: Historical experiment (03/2000-12/2005 mean)

Metrics perform similarly with different model experiments (weights are robust)



Future work: Metrics will be tested on future trends using a "perfect model" approach

- In lieu of having future observations, one model can be treated as the "perfect model" to create weights
- Approach will be repeated choosing different models as the "perfect model" (way to test sensitivity/robustness of metrics to choice of observational data)

Future work: project goals and exploration questions

- Test a variety of weighting metrics for different quantities and regions (Which process metrics are most important for constraining different quantities? Are some metrics better for certain regions?)
- Test scale- and state-dependency of weights (i.e. Do the same models perform consistently better?)
- Apply weights to future projections to construct new ("intelligent") climate predictions